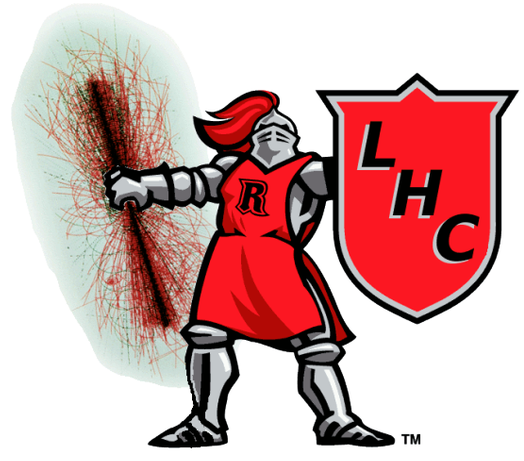
Challenges and Opportunities for Anomaly Detection

David Shih
Rutgers University

Snowmass Community Planning Meeting

Collider Data Analysis Strategies

October 6, 2020



σ(Xtq) x B(X → tW), LH
σ(Btq) x B(B → tW), RH
σ(Bbq) x B(B → tW), RH

B2G-17-01
B2G-17-01
B2G-17-01

Ve

→ tW(LH)
→ tW(RH)
W) = 100%
BB doublet
BB singlet
H) = 100%
Z) = 100%
W) = 100%
TT doublet
TT singlet
H) = 100%
Z) = 100%
W) = 100%
QQ doublet
QQ singlet
H) = 100%
Z) = 100%
W) = 100%

B2G-12-016
B2G-12-016
B2G-12-01

0.0
0.2

-HEP 2019

W' (all final states) B2G-18-006 0.3
W'→WZ (qq̄qq̄) B2G-18-002 0.5
W'→WZ (qq̄νν) B2G-17-005 1.1
W'→WZ (ℓνqq̄) B2G-16-029 2.7
W'→WZ (qq̄ℓℓ) B2G-17-013 5.0
W'→WH (ℓνbb̄) B2G-17-004 4.3
W'→WH (qq̄ττ) B2G-17-006 7.0
W'→WH (qq̄bb̄) B2G-17-002 9.4
Z' (all final states) B2G-18-006 0.6
Z'→WW (qq̄qq̄) B2G-18-002 0.4
Z'→ZH (ℓℓbb̄ + νν bb̄) B2G-17-004 6.0
Z'→ZH (qq̄bb̄) B2G-17-002 6.8
Z'→ZH (qq̄ττ) B2G-17-006 25.0
HVT (all final states) B2G-18-006 0.3
HVT→WW+WZ (qq̄qq̄) B2G-18-002 0.6
HVT→WH+ZH (qq̄bb̄) B2G-17-002 12.0
HVT→WH+ZH (ℓℓbb̄+ℓνbb̄+νν bb̄) B2G-17-004 5.1
HVT→WH+ZH (qq̄ττ) B2G-17-006 7.0
Bulk Graviton k̃=0.5 (all final states) B2G-18-006 2.0
Bulk G→WW+ZZ (qq̄qq̄) B2G-18-002 0.5
Bulk G→WW (ℓνqq̄) B2G-16-029 0.5
Bulk G→ZZ (qq̄νν) B2G-17-005 0.5
Bulk G→ZZ (qq̄ℓℓ) B2G-17-013 1.5
Bulk G→ZZ (ℓℓνν) B2G-16-023 4.0
Bulk G→HH (bb̄bb̄) B2G-16-026 4.4
Radion (Λ_R = 3 TeV)→HH (bb̄bb̄) B2G-16-026 4.4
Radion (Λ_R = 3 TeV)→HH (ℓνqq̄bb̄) B2G-18-008 20.0

HVT model B

0
1.5
3
4.5

CMS, EPS-HEP 2019

95% CL Lower Mass Limit [TeV]
(Upper Cross Section Limit [fb])

Selection of observed limits at 95% C.L.
The quantities ΔM and x represent the
sparticle and the LSP relative to ΔM, re

Very Heav

BB→($\ell^\pm,\ell^\pm\ell^\pm$
BB→($\ell^\pm,\ell^\pm\ell^\pm$
$X_{5/3}X_{5/3}$→tWtW
$X_{5/3}X_{5/3}$→tWtW→($\ell^-,\ell^-\ell^-$) + jets, B($X_{5/3}$→tW)=100%, LH
T_{RH}→tZ→$bq\bar{q}\ell^+\ell^-$, narrow T
bT_{LH}→btZ→$bbq\bar{q}\ell^+\ell^-$, narrow T
B→bH→$b\bar{b}b$, narrow B

$M_{X_{5/3}}$ B2G-17-014 (($\ell^-,\ell^-\ell^-$) + jets) 1.3
$M_{T_{RH}}$ B2G-17-007 (**bqq̄ℓ⁺ℓ⁻**) 1.7
$M_{T_{LH}}$ B2G-17-007 (**bbqq̄ℓ⁺ℓ⁻**) 1.2
M_B B2G-17-009 (**bb̄b**) 1.8

By now, hundreds (thousands?) of searches for new physics at the LHC.

No sign of new physics yet…

CMS, EPS-HEP 2019

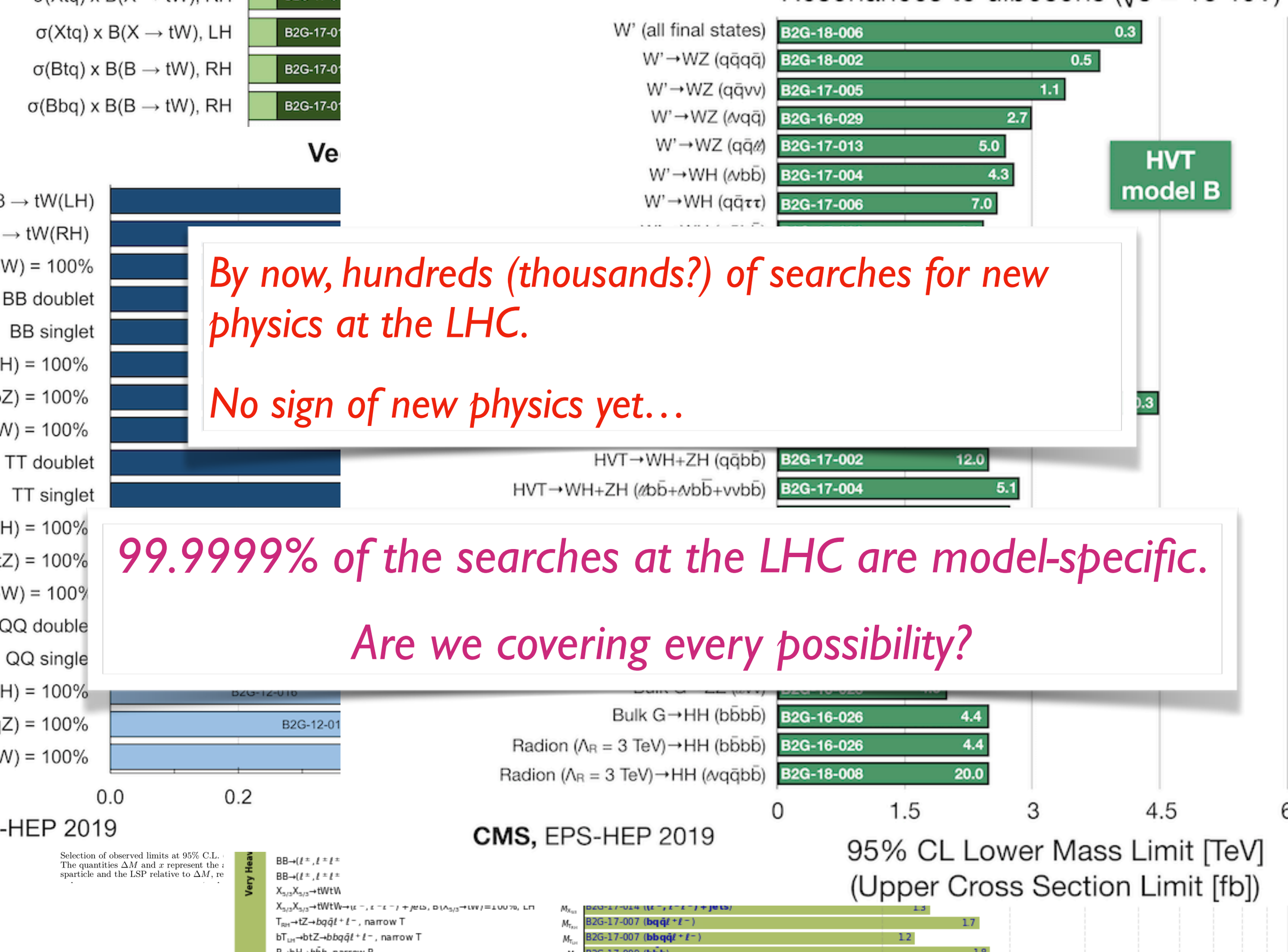
95% CL Lower Mass Limit [TeV]
(Upper Cross Section Limit [fb])

By now, hundreds (thousands?) of searches for new physics at the LHC.

No sign of new physics yet…

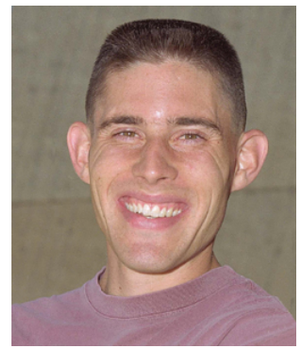
99.9999% of the searches at the LHC are model-specific.

Are we covering every possibility?



Selection of observed limits at 95% C.L.
The quantities ΔM and x represent the
sparticle and the LSP relative to ΔM, re

History of Model Independent Searches at Colliders



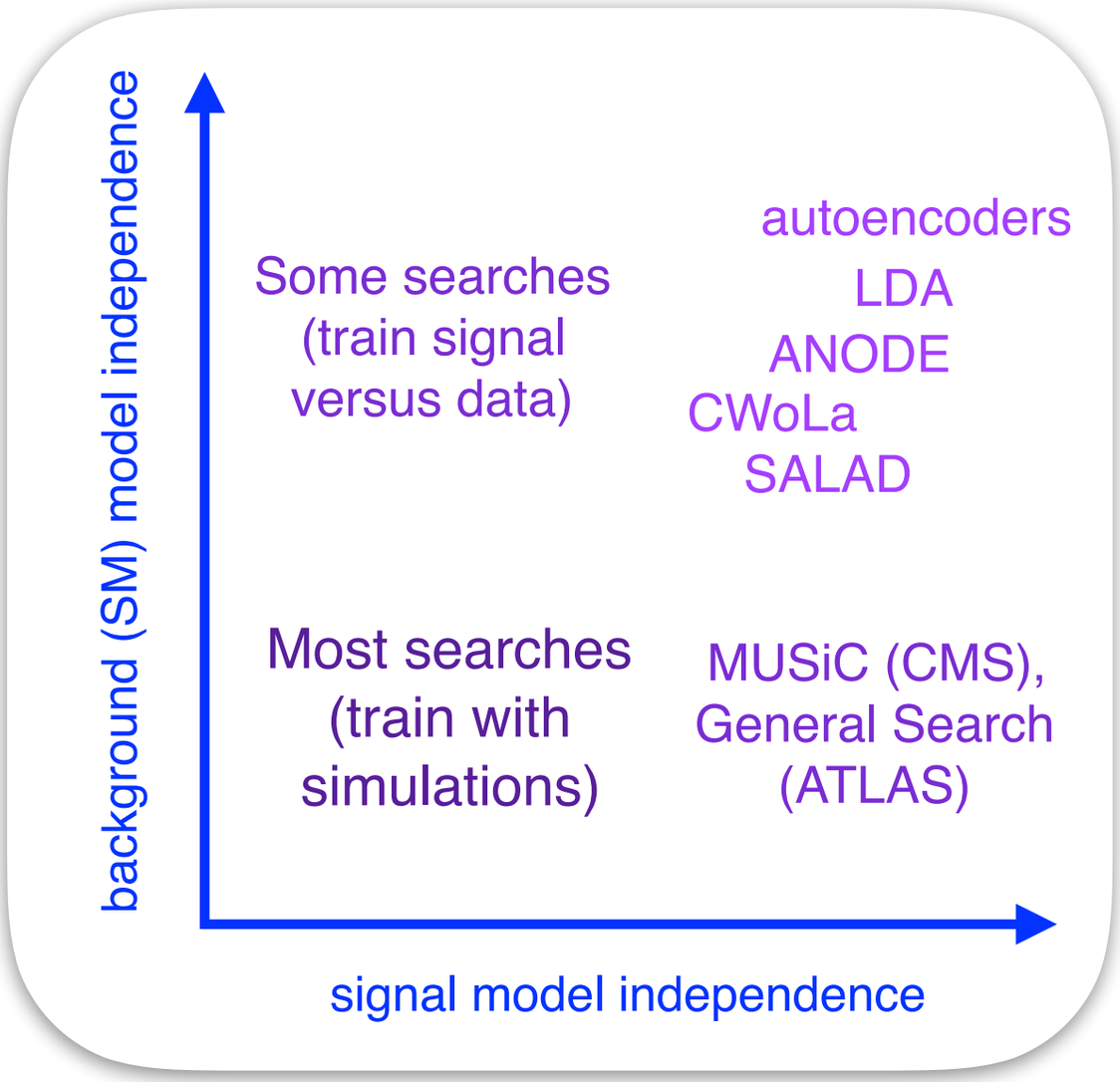
• **D0**	*"Sleuth"*	PRD 62:092004 (2000) PRD 64:012004 (2001) PRL 86:3712 (2001)
• **H1 (Hera)**	*"General Search"*	PLB 602:14-30 (2004) 0705.3721
• **CDF**	*"Sleuth/Vista"*	0712.1311 PRD 78:012002 (2008) 0712.2534 (submitted to PRL, NEVER PUBLISHED) 0809.3781 PRD 79:011101 (2009)
• **CMS**	*"MUSIC"*	CMS-PAS-EXO-14-016
• **ATLAS**	*"Model independent general search"*	1807.07447 EPJC 79:120 (2019)

History of Model Independent Searches at Colliders

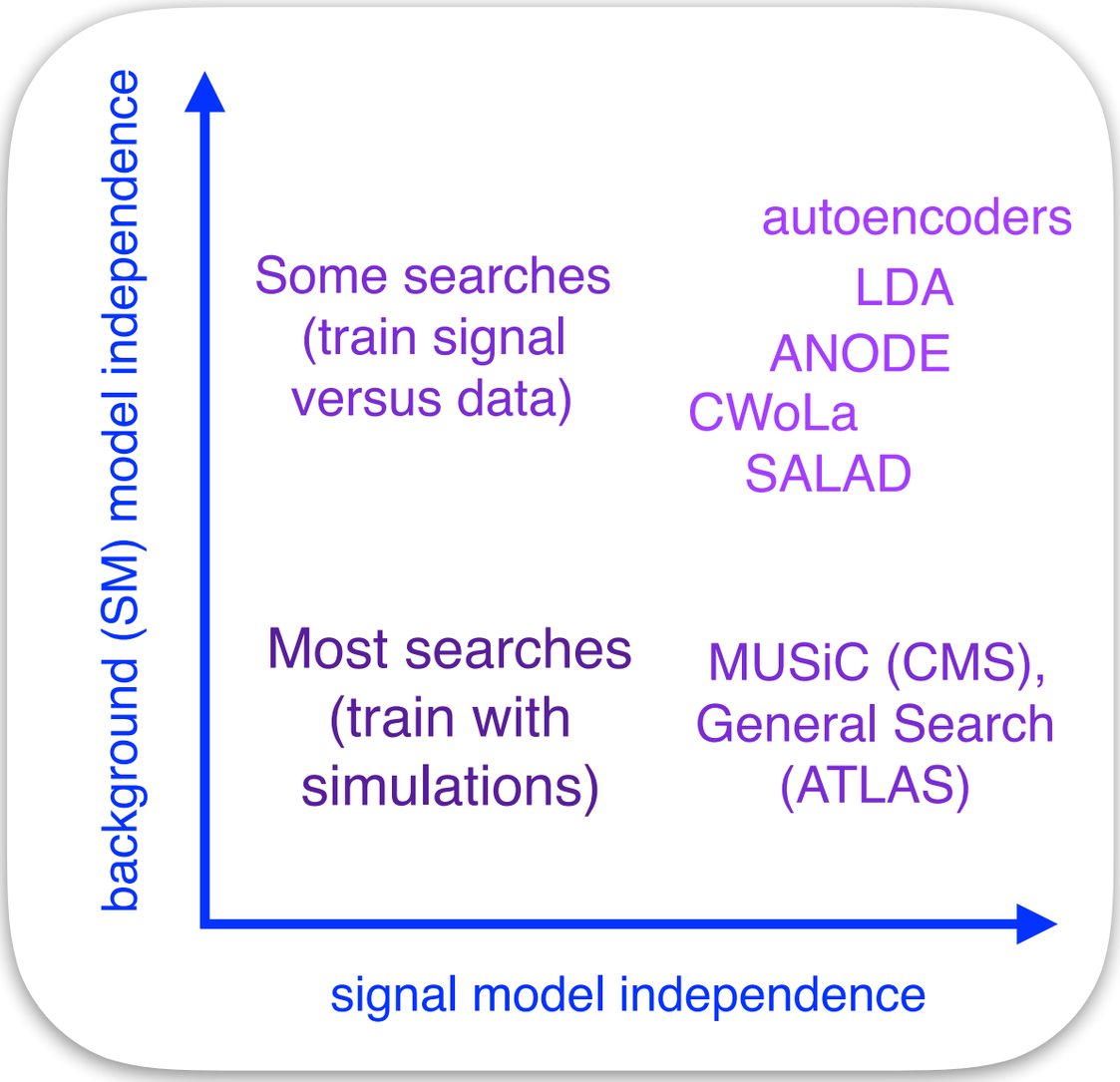
- **D0** *"Sleuth"* PRD 62:092004 (2000) PRD 64:012004 (2001)
- **CMS** *"MUSIC"* CMS-PAS-EXO-14-016
- **ATLAS** *"Model independent general search"* 1807.07447 EPJC 79:120 (2019)

Why are there so few model independent searches at collider experiments?

Why are they all following the same approach — comparing data to simulation?

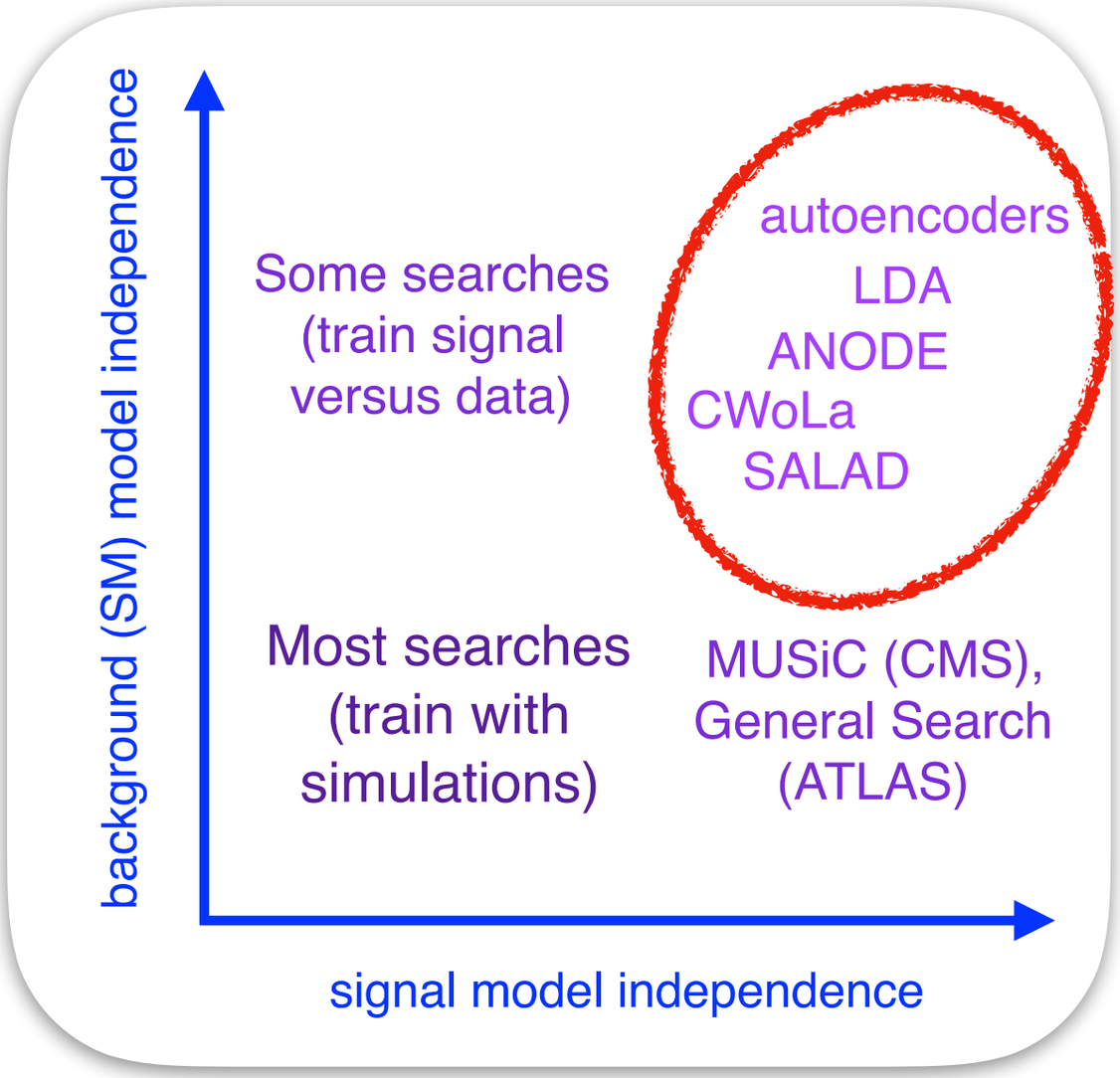


from Nachman & DS 2001.04990



from Nachman & DS 2001.04990

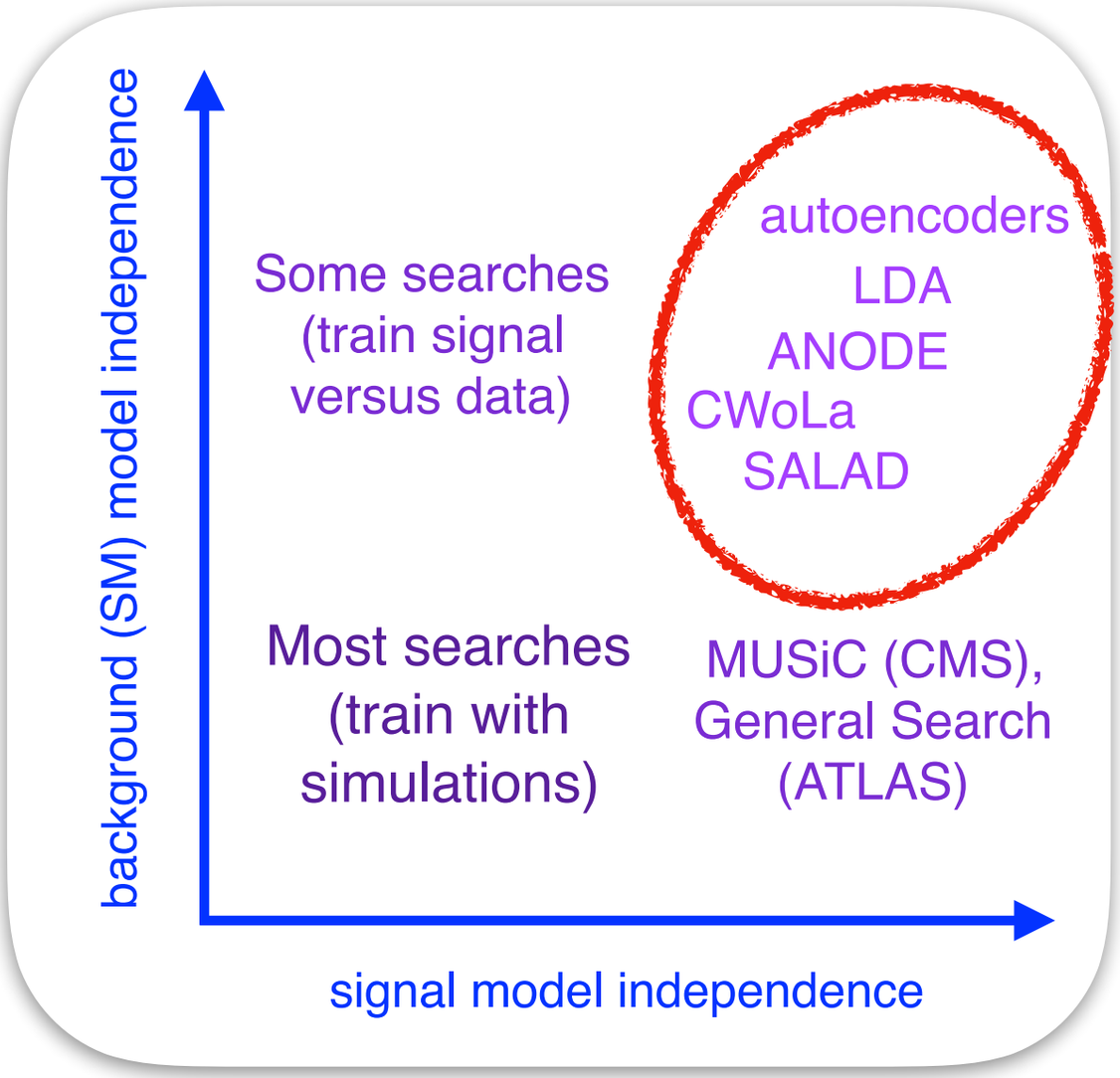
There have been many breakthroughs in unsupervised ML since 2000.



from Nachman & DS 2001.04990

There have been many breakthroughs in unsupervised ML since 2000.

These have inspired many new ideas for model-independent searches at the LHC recently!



from Nachman & DS 2001.04990

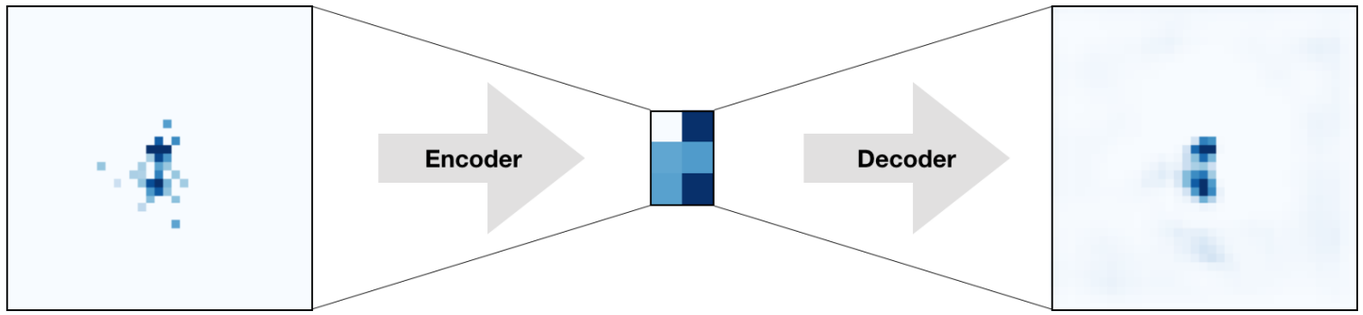
There have been many breakthroughs in unsupervised ML since 2000.

These have inspired many new ideas for model-independent searches at the LHC recently!

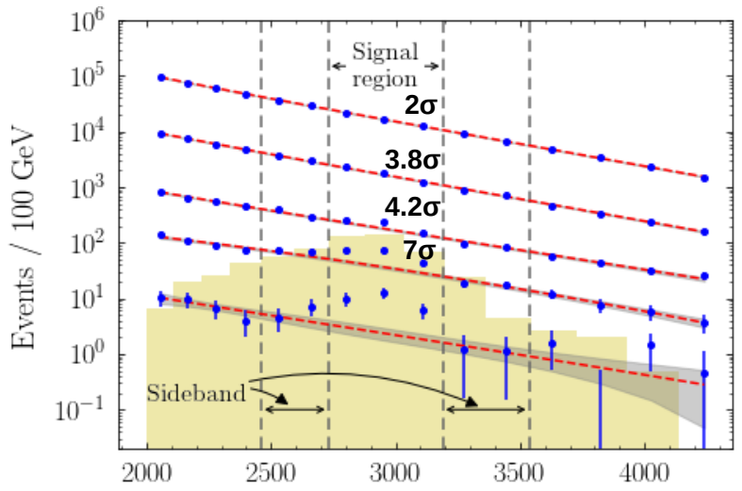
Some challenges:

- Robust background estimation
- Quantifying performance
- Recasting?

Autoencoders

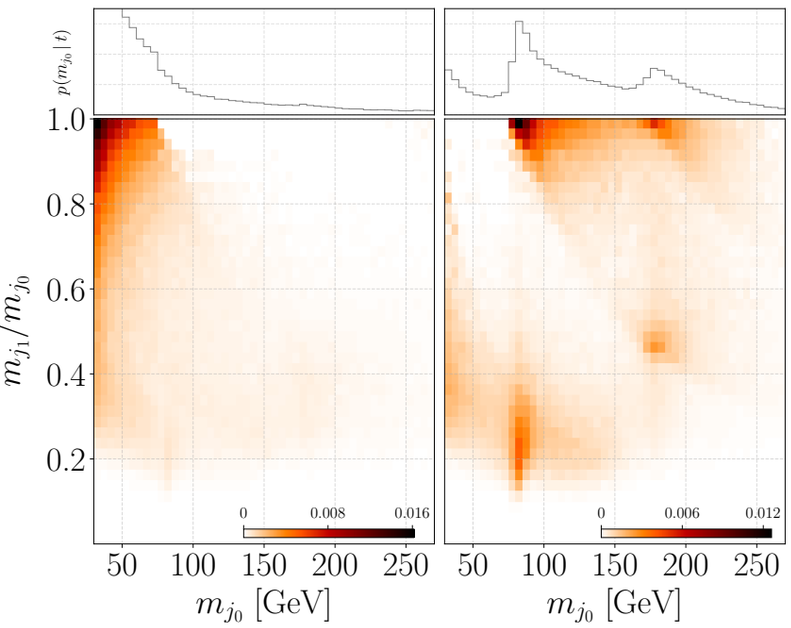


CWoLa

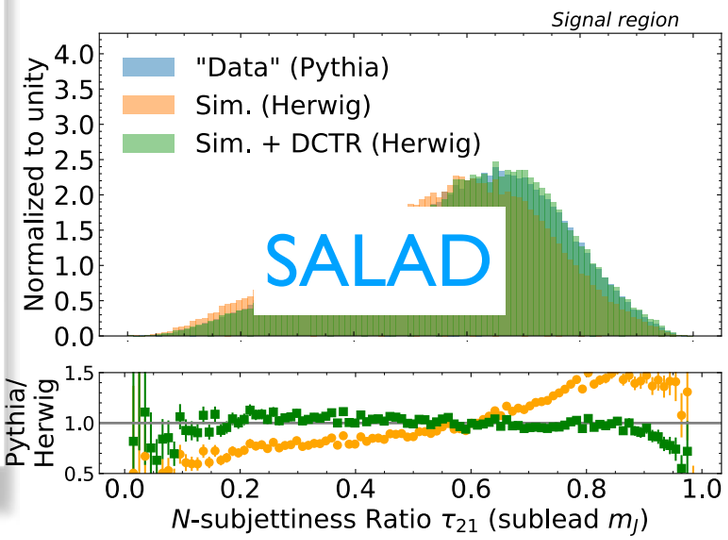


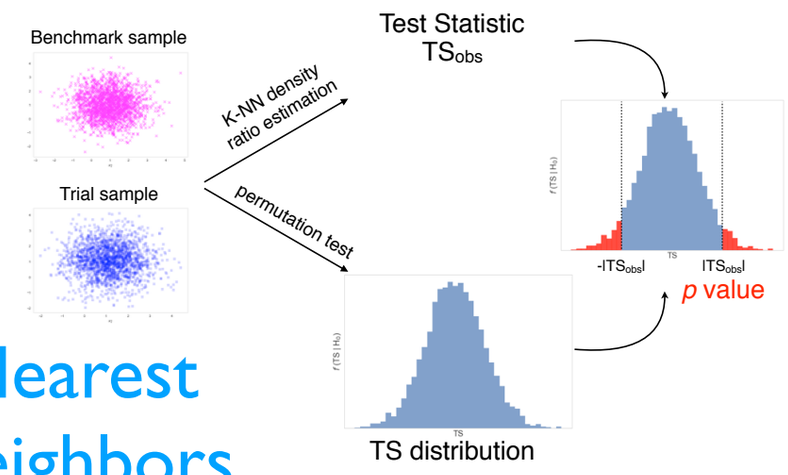
In signal region:
S = 522,
S/B = 0.64%

Probabilistic Modeling

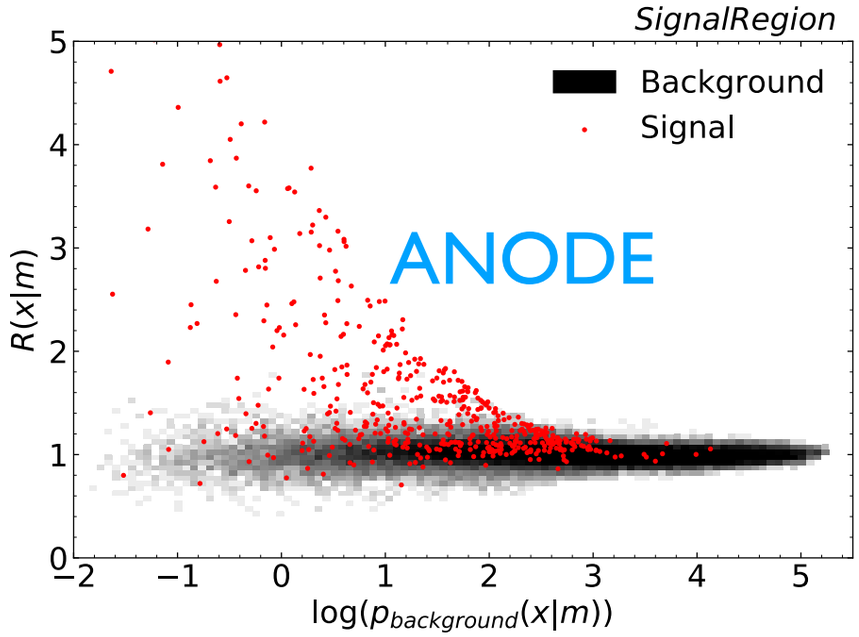


This is just the beginning!





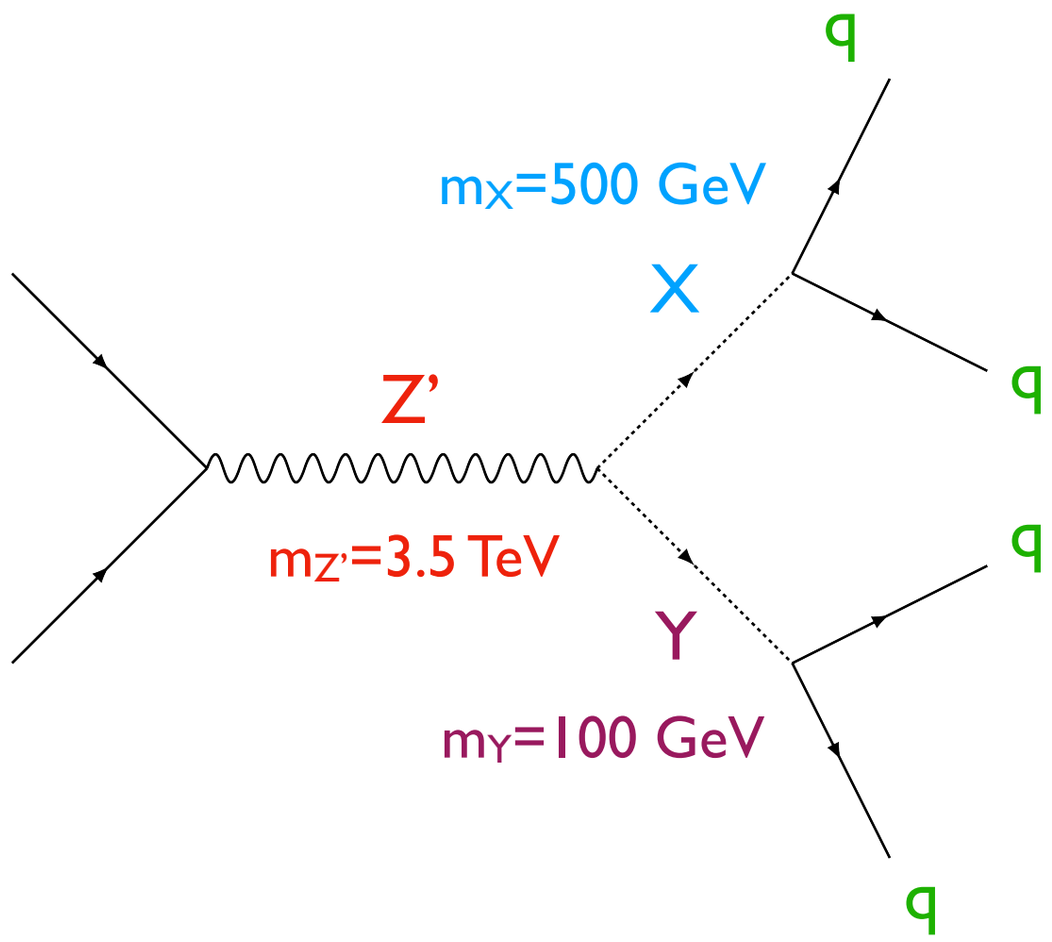
Nearest Neighbors



Backup

A Benchmark Example

LHC Olympics 2020 R&D Dataset
https://doi.org/10.5281/zenodo.2629072



No explicit search at the LHC for this scenario.

Could be hiding in the dijet resonance search at >5sigma significance!!